Bayesian Analysis (0000) **00**, Number 0, pp. 1

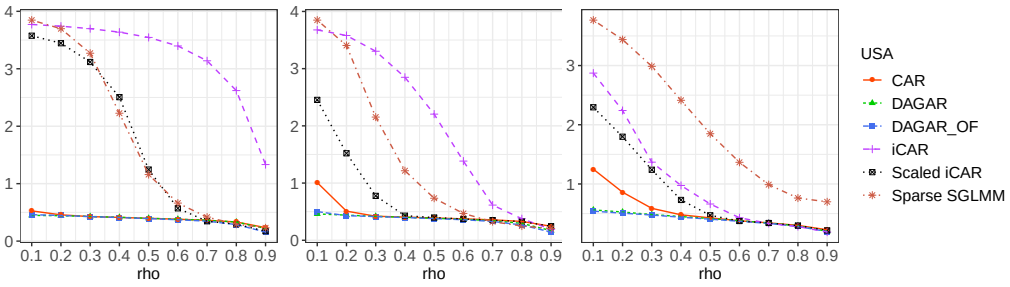
Supplement to "Spatial disease mapping using directed acyclic graph auto-regressive (DAGAR) models"

Abhirup Datta*, Sudipto Banerjee§, James S. Hodges† and Leiwen Gao§

S1 Additional simulation analyses

S1.1 Data generated using a smoother Gaussian Process

As pointed out by one reviewer, in the simulation settings of Section 3.1, the data generation model using an exponential GP becomes same as the DAGAR model for the path graph. While this is not true for the grid and USA graphs, and the results were generally consistent across the choice of the graphs, in this section we tried a different data generation model to assess the performance of the areal models. Keeping all other model specifications same, we generated the spatial random effects w_i using a smoother Matérn$_{3/2}$ GP instead of an exponential GP. This ensures that the data generation model does not correspond to any of the six models fitted to the data for any of the three graphs.



(a) Path (b) Grid (c) USA

Figure S1: MSE as a function of the true ρ (x-axis) for the simulation data analysis using data generated from a Matérn$_{3/2}$ GP

We first look at the mean square error in terms of estimating the latent spatial random effects in Figure S1. We see similar trends as in the case of exponential GP. The MSEs from the ICAR models are much higher, with the scaled ICAR, once again,

*Johns Hopkins University abhidatta@jhu.edu
†University of Minnesota hodge003@umn.edu
§University of California Los Angeles sudipto@ucla.edu
§University of California Los Angeles gaoleiwen@ucla.edu

 DOI: 0000

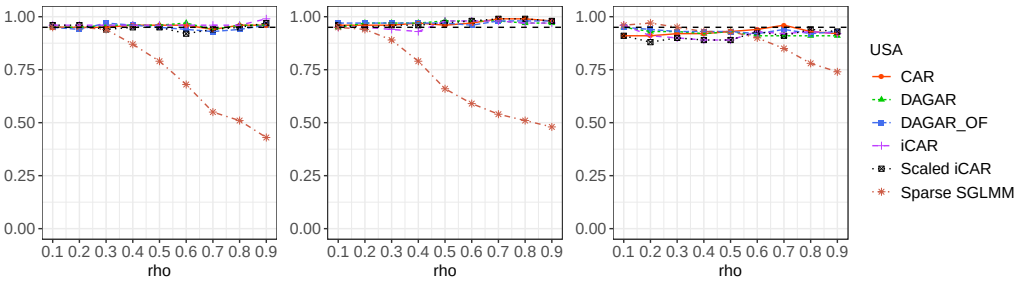
producing lower MSE than the original ICAR and sparse GLMM. The sparse GLMM was better than the ICAR for path and grid graph but was worse for the USA graph. The proper CAR and the two DAGAR models produced lower MSEs than these ICAR-based models for all three graphs, with the improvement more prominent for smaller ρ. For smaller ρ, we also see that the DAGAR models produce lowest MSE among all the six models, whereas for larger ρ, the MSEs for most of the models are similar.

We also briefly summarize the comparison of the models based on inference (CP) on the parameters involved. We only look at the common parameters β_1, β_2 and σ^2_e. We do not consider ρ as, unlike the exponential GP, the spatial decay parameter in the Matérn$_{3/2}$ GP does not have a simple relationship with ρ. Figure S2 provides the coverage probabilities of the three parameters as a function of ρ. We see once again that the trends observed for the exponential GP data analysis in Section 3.1 carry over to here. The coverages for the regression coefficients are close to 95% for all the models except the sparse GLMM. For σ^2_e, all models produce under-coverage for larger values of ρ. For smaller values of ρ, however, the coverage of the proper CAR and the two DAGAR models are close to 95%.

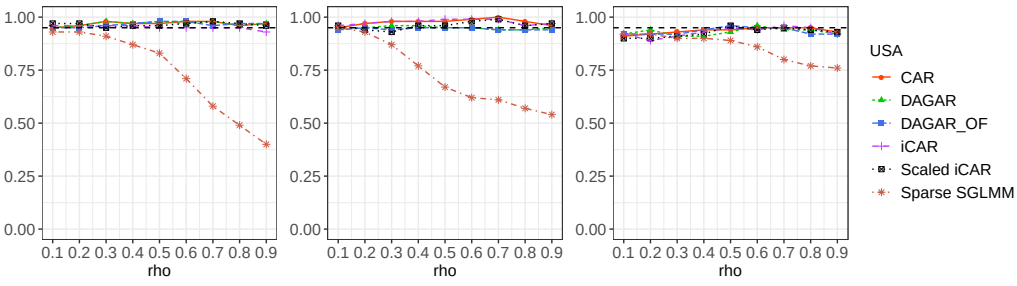
S1.2 A non-Gaussian example

In this Section, we conduct a simulation study using a non-Gaussian response. We generate independent $y_i \sim Poisson(\exp(x_i^\top \beta + w_i))$ where the spatial random effect vector $w = (w_1, w_2, \ldots, w_k)^\top$ are generated as realizations from an exponential GP, akin to Section 3.1. We generated x_i comprising two independent normal variables with mean 0 and standard deviation 0.1 which limits exponential values in regression within the maximum floating point double number in R. Settings for other parameters, τ_w and ρ, are the same as Gaussian data. The priors also remain the same for all models only excluding τ_e which is not included in the model. The same set of six candidate models are fitted.

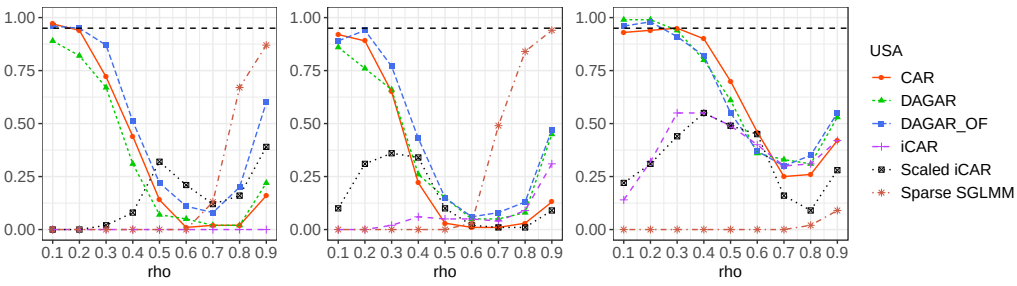
We first compare the MSEs which are quite close for all the models except for the sparse GLMM and ICAR (for path graph) which produce significantly higher MSEs. The DAGAR models produced the lowest MSEs for USA graph, and joint lowest MSEs along-with the proper CAR model for the path graph. We then compare the estimation of ρ for the DAGAR and proper CAR models, as for an exponential GP, ρ corresponds to the correlation at unit distance and the data generation ensured that on average the neighboring units are separated by unit distance (see Section 3.1). The estimates and confidence bands in Figure S4 demonstrates how the DAGAR model produces accurately estimates the spatial correlation between neighbors even when the data is non-Gaussian, whereas the estimates from the CAR model are far off akin to the Gaussian case. Similarly. the coverage probabilities of parameters in Figure S5, repeat the trends observed in Figure 5 for the Gaussian case, with all models except the sparse GLMM offering close to 95% coverage for the regression coefficients, and the DAGAR models offering substantially improved coverage for ρ than the proper CAR model.



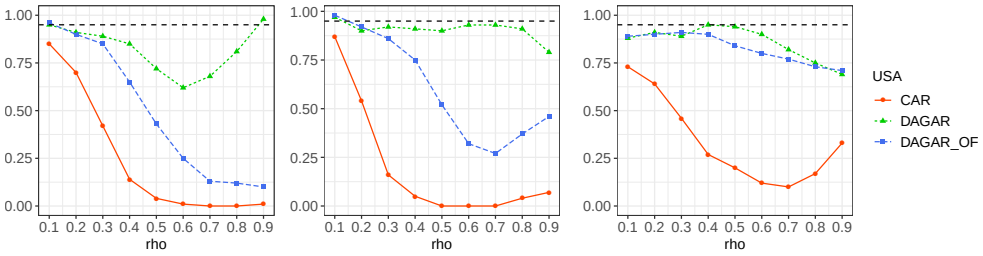
(a) Path: β_1 (b) Grid: β_1 (c) USA: β_1



(d) Path: β_2 (e) Grid: β_2 (f) USA: β_2

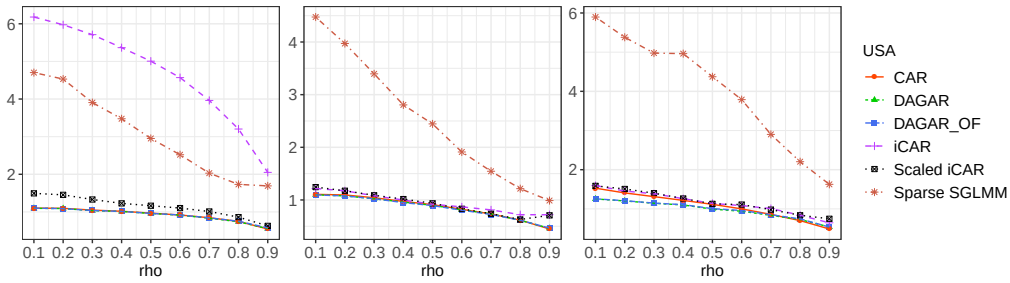


(g) Path: σ_e^2 (h) Grid: σ_e^2 (i) USA: σ_e^2



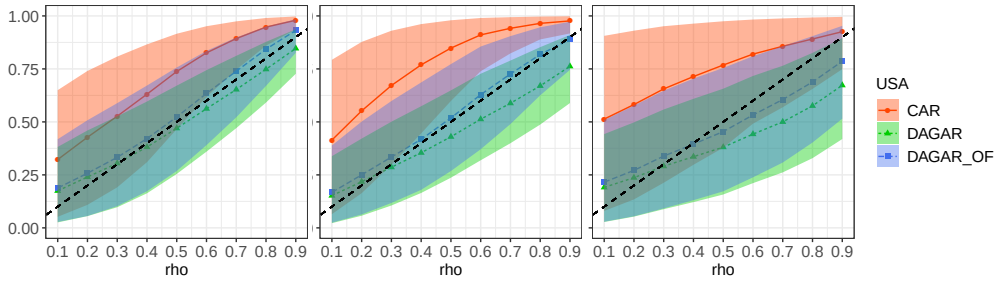
(j) Path: ρ (k) Grid: ρ (l) USA: ρ

Figure S2: Coverage probabilities of the parameters as a function of the true ρ (x-axis) for the simulation data analysis using Gaussian data generated from a Matérn$_{3/2}$ GP



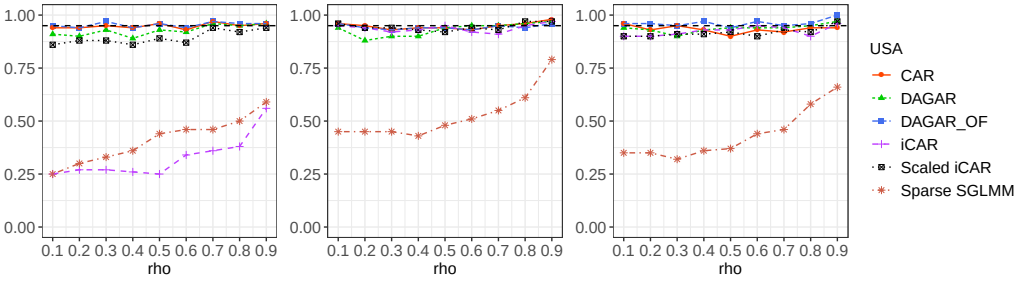
(a) Path (b) Grid (c) USA

Figure S3: MSE as a function of the true ρ (x-axis) for the simulation data analysis using Poisson responses

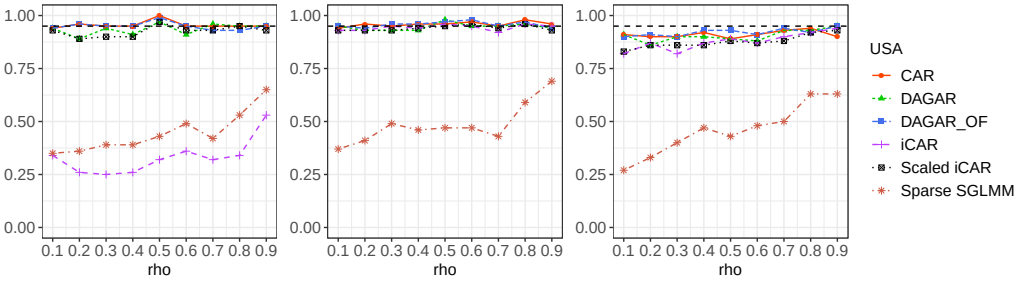


(a) Path (b) Grid (c) USA

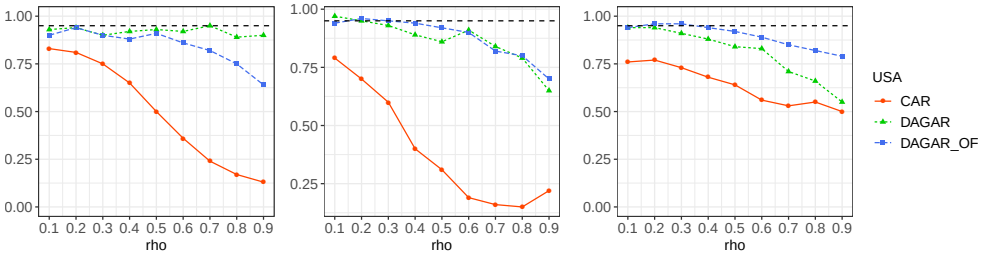
Figure S4: Estimate and confidence bands of ρ as a function of the true ρ (x-axis) for the simulation data analysis using Poisson responses



(a) Path: β_1 (b) Grid: β_1 (c) USA: β_1



(d) Path: β_2 (e) Grid: β_2 (f) USA: β_2



(g) Path: ρ (h) Grid: ρ (i) USA: ρ

Figure S5: Coverage probabilities of the parameters as a function of the true ρ (x-axis) for the simulation data analysis using Poisson responses

S2 Proofs

S2.1 Proof of Theorem 1

Let $r = rank(Q)$ and Q^+ denote the Moore-Penrose inverse of Q. Then, by Theorem 1 part (b) of Higham (1990), there exists a permutation P such that

$$PQ^+P^\top = R^\top R \text{ where } R = \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix}$$

Here R_1 is $r \times r$ upper triangular matrix with positive diagonal elements. Let $D_1 = diag(R_1)$ and $R_1^* = D_1^{-1}R_1$, which has ones on the diagonal. We can now write

$$R = DU \text{ where } U = \begin{bmatrix} R_1^* & D_1^{-1}R_2 \\ 0 & I \end{bmatrix} \text{ and } D = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix}$$

Since $U^\top$ is a lower triangular matrix with one on the diagonals, so is $L = U^{-\top}$. Hence, $PQP^\top = (I-B)^\top F(I-B)$ where $F = D^{+2}$ and $B = I - L$ is a strictly lower triangular matrix.

S2.2 Proof of Theorem 2

First of all, as $\mathcal{T}$ is a tree, it is always possible to have an ordering π such that $n_\pi(i) = 1$ for any $i \neq \pi(1)$. For example, the orderings corresponding to any pre-order or breadth-first tree traversal of $\mathcal{T}$ will satisfy this. Without loss of generality we rename the nodes such that $\pi = \{1, \ldots, k\}$ and for $i > 1$, $p(i)$ denotes the directed neighbor of i in π implying $p(i) < i$. Letting $w_0 = 0$, $p(1) = 0$ and $n_\pi(1) = 0$, the model in (2.5) reduces to $w_i = \rho\ w_{p(i)} + (1-\rho^2)^{0.5}\epsilon_i$ where ϵ_i are independent standard normal variables. We shall show that for any positive integers $j \leq i \leq k$, $cov(w_i, w_j) = \rho^{d_{ij}}$. We prove this using the strong form of mathematical induction. Since $p(2) = 1$, it is easy to verify this for $i = 2$. We assume that this is true for $i = 2, \ldots, i-1$. It immediately follows that $var(w_i) = \rho^2 var(w_{p(i)}) + (1-\rho^2)var(\epsilon_i) = 1$. For any $j < i$, $cov(w_i, w_j) = \rho\ cov(w_{p(i)}, w_j) = \rho^{1+d_{p(i)j}}$ (by induction). Since $\mathcal{T}$ is acyclic and $n_\pi(i) = 1$ for all $i > 1$, the shortest path from j to i runs through $p(i)$. Hence, $d_{ij} = d_{p(i)j} + 1$ and the result follows.

S2.3 Proof of Theorem 3

If $i + j = i' + j'$, then (i, j) and (i', j') are never neighbors. Hence, without loss of generality, we prove the result for $\pi = (S_2, \ldots, S_{m+n})^T$ where $S_r = \{(i, j) \,|\, i + j = r\}$. Let $d((i, j), (i', j')) = |i - i'| + |j - j'|$ denote the Manhattan distance on $\mathcal{G}$, and w_{S_r} denote the sub-vector of w corresponding to the indices in S_r. It is enough to show by induction on r that $w_{S_r} \sim N(0, (\rho^{D_r})^{-1})$ where D_r denotes the distance matrix on S_r. This holds trivially for $r = 2$. Let us assume that it holds true for $r-1$. If $(1, r-1) \in S_r$, we define $w(0, r-1) = \rho w(1, r-1) + \epsilon(0, r-1)$ where $\epsilon(0, r-1) \sim N(0, 1/(1-\rho^2))$ is independent of w. If $(r-1, 1) \in S_r$ we define $w(r-1, 0)$ similarly. Let $w^*_{S_{r-1}}$ be the

augmented vector which includes $w(0, r-1)$ or $w(r-1, 0)$ or both, along with $w_{S_{r-1}}$. From the construction, $w(0, r-1) = \rho^2 w(1, r-2) + \rho\epsilon(1, r-1) + \epsilon(0, r-1)$ implying $var(w(0, r-1)) = 1$ and $cov(w(0, r-1), w(1, r-2)) = \rho^2$. Hence $cov(w^*_{S_{r-1}}) = \rho^{D^*_r}$ where D^*_r is the augmented distance matrix corresponding to S^*_{r-1}. Letting $\rho^2 = u$, we have for any (i, j) and (i', j') in S_r,

$$\begin{aligned} cov(w(i,j),w(i',j')) &= \frac{u}{(1+u)^2}\left(cov(w(i-1,j),w(i'-1,j')) + cov(w(i,j-1),w(i'-1,j'))\right. \\ &\quad \left. + cov(w(i-1,j),w(i',j'-1)) + cov(w(i,j-1),w(i',j'-1))\right) + I(i=i')\frac{1-u}{1+u} \\ &= \frac{u}{(1+u)^2}(\rho^{|i-i'+1|+|j-j'-1|} + 2\rho^{|i-i'|+|j-j'|} + \rho^{|i-i'-1|+|j-j'+1|}) + I(i=i')\frac{1-u}{1+u} \end{aligned}$$

If $i = i'$ then $j = j'$ and the expression above equals 1. If $i < i'$, then $j > j'$ and $|i - i' + 1| + |j - j' - 1| = (i' - i - 1) + (j - j' - 1) = |i - i'| + |j - j'| - 2$. Similarly, $|i - i' - 1| + |j - j' + 1 = |i - i'| + |j - j'| + 2$. So, $\rho^{|i-i'+1|+|j-j'-1|} + 2\rho^{|i-i'|+|j-j'|} + \rho^{|i-i'-1|+|j-j'+1|} = \rho^{|i-i'|+|j-j'|}(1/u + 2 + u)$. Hence, the results follows.

S2.4 Proof of Theorem 4

For any vertex i with n_i neighbors, let π_{ir} denote the set of all permutations π such that $n_{\pi(i)} = r$. By symmetry, $|\pi_{ir}| = k!/(n_i + 1)$ for $r = 0, 1, \ldots, n_i$. Also, for any $i \sim j$ and $r = 0, \ldots, n_i$, let π_{ijr} denote the set of all permutations π such that $n_{\pi(i)} = r$ and $j \in N_\pi(i)$. Then, $|\pi_{ijr}| = k!/(n_i + 1) \times pr(j \text{ is among the } r \text{ directed neighbors of } i) = rk!/(n_i(n_i + 1))$. We now have

$$\begin{aligned} Q[i,i] &= \frac{1}{k!(1-\rho^2)}\sum_\pi \left(1 + (n_{\pi(i)} - 1)\rho^2 + \sum_{j\sim i} I(i \in N_\pi(j))\frac{\rho^2}{1 + (n_{\pi(j)} - 1)\rho^2}\right) \\ &= 1 + \frac{\rho^2}{k!(1-\rho^2)}\left(\sum_{r=0}^{n_i} r|\pi_r| + \sum_{j\sim i}\sum_{r=0}^{n_j}\frac{|\pi_{jir}|}{1 + (r-1)\rho^2}\right) \\ &= 1 + \frac{n_i\rho^2}{2(1-\rho^2)} + \frac{\rho^2}{1-\rho^2}\sum_{j\sim i}\frac{1}{n_j(n_j+1)}\sum_{r=1}^{n_j}\frac{r}{1 + (r-1)\rho^2} \\ &= 1 + \frac{n_i\rho^2}{2(1-\rho^2)} + \frac{\rho^2}{1-\rho^2}\sum_{j\sim i}\frac{1}{n_j(n_j+1)} f(\rho, n_j). \end{aligned}$$

To evaluate the non-diagonal entries of Q, we additionally define π_{ijkr} to be the set of all permutations π such that $n_{\pi(i)} = r$ and $\{j, k\} \subseteq N_\pi(i)$. Applying the combinatorial argument used earlier, we see that $|\pi_{ijkr}| = r(r-1)k!/((n_i - 1)n_i(n_i + 1))$. Let $i \approx j$ implies that there exists at least one node k such that $i \sim k$ and $j \sim k$.

$$Q[i,j] = \frac{1}{k!(1-\rho^2)}\sum_\pi \left(-\rho I(i \sim j) + I(i \approx j)\sum_{k:\{i,j\}\subseteq N_\pi(k)}\frac{\rho^2}{1 + (n_{\pi(k)} - 1)\rho^2}\right)$$

$$= -\frac{\rho}{1-\rho^2} I(i \sim j) + \frac{\rho^2}{k!(1-\rho^2)} I(i \approx j) \sum_{k \sim N(i) \cap N(j)} \sum_{r=0}^{n_k} \frac{|\pi_{kijr}|}{1+(r-1)\rho^2}$$

$$= -\frac{\rho}{1-\rho^2} I(i \sim j) + \frac{\rho^2}{1-\rho^2} I(i \approx j) \sum_{k \sim N(i) \cap N(j)} \frac{1}{(n_k-1)n_k(n_k+1)} \sum_{r=1}^{n_k} \frac{r(r-1)}{1+(r-1)\rho^2}$$

$$= -\frac{\rho}{1-\rho^2} I(i \sim j) + \frac{1}{1-\rho^2} I(i \approx j) \sum_{k \sim N(i) \cap N(j)} \left(\frac{1}{2(n_k-1)} - \frac{1}{(n_k-1)n_k(n_k+1)} f(\rho, n_k) \right).$$

S2.5 Proof of Theorem 5

We write $Q_\pi(\rho)$ and $Q(\rho)$ as Q_π and Q hiding the dependence on ρ except when necessary. From Theorem 4 we have for $i = 3, \ldots, k-2$, $Q_{ii} = \frac{3+6\rho^2+\rho^4}{3(1+\rho^2)(1-\rho^2)}$, $Q_{i,i+1} = -\frac{\rho}{1-\rho^2}$ and $Q_{i,i+2} = \frac{\rho^2}{3(1+\rho^2)(1-\rho^2)}$. Hence,

$$||Q||_F^2 = \frac{k}{9(1+\rho^2)^2(1-\rho^2)^2} \left((3+6\rho^2+\rho^4)^2 + 18\rho^2(1+\rho^2)^2 + 2\rho^4 \right) + o(k)$$

where the $o(k)$ term arises from rows and columns corresponding to the nodes at the extreme right or left.

Now using left-to-right or right-to-left ordering, from (2.4), a typical term in the quadratic form $w'Q_\pi w$ will be of the form $(w_i - \rho w_{i-1})^2/(1-\rho^2)$. Hence, for $i = 1, 2, \ldots, k-2$, $Q_{\pi:ii} = \frac{1+\rho^2}{1-\rho^2}$, $Q_{\pi:i,i+1} = -\frac{\rho}{1-\rho^2}$ and $Q_{\pi:i,i+2} = 0$. So,

$$\begin{aligned} ||Q - Q_\pi||_F^2 &= \frac{k}{9(1+\rho^2)^2(1-\rho^2)^2} \left((3+6\rho^2+\rho^4 - 3(1+\rho^2)^2)^2 + 2\rho^4 \right) + o(k) \\ &= \frac{k}{9(1+\rho^2)^2(1-\rho^2)^2} (4\rho^8 + 2\rho^4) + o(k) \end{aligned}$$

Hence, the result follows.

S2.6 Proof of Theorem 6

We index the nodes of the grid as (i, j), and entries of Q_π and Q as $Q_{\pi:(ij),(i'j')}$ and $Q_{(ij),(i'j')}$ respectively for $1 \le i, i', j, j' \le m$. Like in the proof of Theorem 5, it only suffices to evaluate the Frobenius norms for the interior points of the grid (having 4 neighbors each of whom also have 4 neighbors) as the contribution from the remaining terms will be $o(m^2)$. Hence from Theorem 3 we have for $3 \le i, i', j, j' \le m-2$,

$$\begin{aligned} Q_{(ij),(ij)} &= \frac{1+\rho^2+\rho^2 s(\rho)/5}{1-\rho^2} \\ Q_{(ij),(i+1,j)} &= Q_{(ij),(i,j+1)} = -\frac{\rho}{1-\rho^2} \\ Q_{(ij),(i+2,j)} &= Q_{(ij),(i,j+2)} = \frac{1}{1-\rho^2}(1/6 - s(\rho)/60) \end{aligned}$$

$$Q_{(ij),(i+1,j+1)} = Q_{(ij),(i+1,j-1)} = \frac{2}{1-\rho^2}(1/6 - s(\rho)/60)$$

Summing up, we have

$$\begin{aligned}||Q||_F^2 &= m^2(Q_{(ij),(ij)}^2 + 4(Q_{(ij),(i+1,j)}^2 + Q_{(ij),(i+2,j)}^2 + Q_{(ij),(i+1,j+1)}^2) + o(1)) \\ &= \frac{m^2}{(1-\rho^2)^2}\left((1+\rho^2+\rho^2 s(\rho)/5)^2 + 4\rho^2 + 20(1/6 - s(\rho)/60)^2\right) + o(m^2)\end{aligned}$$

Now, without loss of generality we assume that the DAGAR precision matrix $Q_\pi(\rho)$ was constructed by ordering the nodes in increasing order of $(i+j)$. Then, a typical term in the quadratic form $w'Q_\pi w$ will be of the form $\frac{1+\rho^2}{1-\rho^2}(w_{ij} - \frac{\rho}{1+\rho^2}(w_{i,j-1} + w_{i-1,j}))^2$. Hence, we will have

$$\begin{aligned}Q_{\pi:(ij),(ij)} &= \frac{1+\rho^2+2\rho^2/(1+\rho^2)}{1-\rho^2} \\ Q_{\pi:(ij),(i+1,j)} &= Q_{\pi:(ij),(i,j+1)} = -\frac{\rho}{1-\rho^2} \\ Q_{\pi:(ij),(i+2,j)} &= Q_{\pi:(ij),(i,j+2)} = Q_{\pi:(ij),(i+1,j+1)} = 0 \\ Q_{\pi:(ij),(i+1,j-1)} &= \frac{\rho^2}{(1+\rho^2)(1-\rho^2)}\end{aligned}$$

Subtracting, we have

$$\begin{aligned}||Q_\pi - Q||_F^2 = m^2((&Q_{\pi:(ij),(ij)} - Q_{(ij),(ij)})^2 + 4Q_{(ij),(i+2,j)}^2 + \\ &2(Q_{\pi:(ij),(i+1,j-1)} - Q_{(ij),(i+1,j-1)})^2 + \\ &2Q_{(ij),(i+1,j+1)}^2 + o(1))\end{aligned}$$

Hence, the result follows.

References

Higham, N. J. (1990). "Analysis of the Cholesky decomposition of a semi-definite matrix." In *in Reliable Numerical Computation*, 161–185. University Press. 6